DOCUMENT RESUME

ED 391 358 FL 023 479

AUTHOR Gu, Yongqi; And Others
TITLE How Often Is Often? Reference Ambiguities of the Likert-Scale in Language Learning Strategy Research.
PUB DATE 95
NOTE 18p.; In: Occasional Papers in English Language Teaching, Volume 5, p19-35, 1995 published by Chinese University of Hong Kong. An earlier version of this paper was presented at the 1993 Guilin English Language Teaching International Conference.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Ambiguity; College Students; *English (Second Language); Foreign Countries; Higher Education; Language Research; Learning Strategies; *Likert Scales; Questionnaires; Rating Scales; Second Language Learning; *Statistical Analysis; Test Format
IDENTIFIERS China

ABSTRACT

Based on personal experience, this paper examines the ambiguities of the Likert-type 5-point scale in learning strategy elicitation. Four parallel questionnaires consisting of the same batch of 20 items taken from the Oxford scale (1990) were administered among a group of 120 tertiary level, non-English majors in China. Questionnaire 1 used the Oxford scale without specifying dimensions of reference. Questionnaire 2 told the respondents to choose their answers by comparing with their peers in the same grade. Questionnaire 3 asked them to select their present behavioral frequency as compared with their own past learning experience in secondary schools. In questionnaire 4, subjects were told to check off the relevant frequency of a behavior by comparing its frequency of occurrence with that of other language skills. Results showed that out of the 20 items used, 13 were significantly different among the four questionnaires. Methodological implications for questionnaire research are discussed, and suggestions for future research are proposed. (Contains 25 references.) (Author/NAV)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Occasional Papers in English Language Teaching Vol. 5, 1995
ELT Unit, Chinese University of Hong Kong. 19-35.

ED 391 358

HOW OFTEN IS *OFTEN*?

REFERENCE AMBIGUITIES OF THE LIKERT-SCALE IN LANGUAGE LEARNING STRATEGY RESEARCH

Yongqi Gu
University of Hong Kong

Qiufang Wen
Nanjing University

Danxia Wu
Central Institute of Fine Arts

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Gu Yongqi

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Studies in second language acquisition, especially in the area of language learning strategies, frequently employ the survey method alone as their means of investigation. Incongruent results are normally explained in terms other than the survey measure as such. One of our recent qualitative studies, however, revealed that respondents have different reference systems in mind when answering Likert-type questions. In this study, we call into question the ambiguities of the Likert-type five-point scale in learning strategy elicitation. Four parallel questionnaires consisting of the same batch of 20 items taken from Oxford (1990) were administered among a group of 120 tertiary level non-English majors in China. Questionnaire 1 directly took Oxford's scale without specifying dimensions of reference; Questionnaire 2 told the respondents to choose their answers by comparing with their peers in the same grade; Questionnaire 3 asked them to select their present behavioral frequency as compared with their own past learning experience in secondary schools; and in Questionnaire 4, subjects were told to tick the relevant frequency of a behavior by comparing its frequency of occurrence with that of other language skills. Data from the four questionnaires were subjected to repeated measures MANOVA analysis using SPSS/PC+. Results showed that out of the 20 items, 13 were significantly different among the four questionnaires. Methodological implications for questionnaire research are next discussed and suggestions for future research proposed.

FL023479

The survey technique that uses an ordinal scale to measure the strength of an attitude, and uses several items to form an attitudinal construct is usually referred to as a Likert-scale (Shaw & Wright, 1967). Since Likert (1932) modified Thurstone's (1928) scaling method and made it an easy-to-use measurement technique, the Likert-scale has flourished for decades in social and behavioral research. By far it is most often applied to attitudinal measurement;

2 BEST COPY AVAILABLE

fewer studies, however, employ the Likert-scale as a yardstick for human behavior (Dunn-Rankin, 1983). When measuring attitudes, it usually takes the form of a five-point scale that indicates a respondent's strength of agreement to a statement (from *strongly disagree* to *strongly agree*). When behaviors are the target of measurement, on the other hand, the scale becomes the measurement of the frequency with which a behavior is thought to occur.

Numerous problems have been reported on the validity and reliability of the Likert scale (see, for example, Busch, 1993; Keppel, 1991; Turner, 1993). Some of these problems result from the scale itself, others from its applications. For instance, one of the widely used formats for the elicitation of behavioral frequency— *never, rarely, sometimes, often, always* — is quite often dubious due to its semantic inexplicitness. Take the word *often*, for example. Different individuals will almost certainly disagree on how frequently an action is to take place before being regarded as *often*. One solution to this problem is to spell out the frequency of occurrence of a behavior. Still, one needs to take meticulous care about how he does the specification simply to avoid even more confusion. As an example, Oxford's (1990) explanation to *somewhat true of me* as *true of me for half of the time* (p. 293) may well be argued to have added more trouble than illumination. "What is '*half of the time*?'" Wen (1993) asked, "Half of the time when I am awake, half of all my time spent on learning, or what?"

Another related problem does not quite lie in the scale as such. It is not unusual to see results from the Likert type questionnaire subjected to a statistical analysis that presumes a linear relationship between the psychological or behavioral construct tested by the scale and a criterion measure when in fact the relationship is other than linear. Granted that simple correlations between each questionnaire item and the dependent variable measure may not greatly distort the actual picture, when a construct resulting from several items averaged is correlated with the dependent measure, distortion is much more likely to occur if the relationship between some items in the construct and the criterion measure is linear while the relationship between the other items in the construct and the same criterion measure is not. Moreover, even if the whole construct does enjoy homogeneity in terms of its relationship with the criterion measure, confusion is still likely to result from more sophisticated statistical tests such as multiple regression, LISREL, or path analysis where all constructs are put together for linear modeling. To be more specific, the relationship between anxiety as measured via Likert-type questionnaires and learning outcome is known to be non-linear, which by no means suggests that anxiety is not important in learning. However, a linear analysis of the two constructs would produce a result suggesting a weak relationship between them. The best way to prevent this from happening is to plot each questionnaire item and each construct against each criterion measure before subjecting them to further analysis.

In addition, response sets—especially cultural differences in response sets, a problem directly associated with Likert scaling—have also been bothering social scientists for a long time (e.g., Hui & Triandis, 1989; Triandis & Triandis, 1962). For example, it has been repeatedly demonstrated that the Asians differ from the British (Wright et al., 1978) and Hispanics (Hui & Triandis, 1989) in terms of what they exactly mean when they respond to Likert-type questions. Zax & Takahashi (1967) also have reported that Asians tend to use the middle of the scale and take it as an indication of their highly valued modesty, whereas Mediterranean people tend to use extreme responses to show their sincerity. While these findings are fully justified, we nevertheless believe that even people from a homogeneous cultural background may also differ in terms of what they really mean when they choose the same answer. In other words, individual respondents may well have very different subjective reference systems when presented with a relative scale.

These problems are particularly relevant to research in SLA, as the bulk of work on language learning strategies, for instance, frequently employs the survey method alone as a means of investigation (e.g., Oxford & Nyikos, 1989; Oxford, Nyikos, & Ehrman, 1988; Politzer & McGroarty, 1985). Incongruent results are normally explained in terms other than the measurement as such (Gu, 1992). One of our recent qualitative studies (Wen, 1993), however, has revealed that respondents' different reference systems might have influenced the ways the Likert-type questions were answered. For example, some subjects complained about not knowing whom to compare with when asked about how often they performed a learning behavior. "What do you mean by *often*?" asked one. "Compared to my classmates, I seldom do it. Compared to myself several years ago in my secondary school, however, I'm doing it quite often." "Compared to listening and reading," said another, "I rarely do any speaking and writing at all." To make matters worse, some subjects reported that they might compare with their classmates when answering one item, and compare with their own past learning experiences when answering another item.

Obviously, these subjective reference variations distort to a considerable extent the interpretation of survey results, so much so, in fact, that we began to doubt the reliability of any general survey measure that relies solely upon the Likert-scale as its indicator of learning behaviors short of backing it up with other means of data elicitation. The present study was thus designed to confirm or reject our doubt and to see whether different questionnaires that specify different systems of reference would yield different results.

DESIGN

Subjects

The participants in this study were 120 sophomore science students who were learning English as a foreign language at Beijing University of Industry. Of the 120, 110 students returned the questionnaires. A brief examination of the returned questionnaires reduced the usable number of subjects to 95. For example, subjects who chose the same answer for the whole questionnaire were eliminated from the data pool, and subjects who did not fill in their names were also excluded from final analysis.

Materials

Four parallel questionnaires consisting of the same batch of 20 items taken or adapted from Oxford (1990) were administered. Questionnaire 1 directly took Oxford's scale without specifyin9g dimensions of reference; Questionnaire 2 told the respondents to choose their answers by comparing with their peers; Questionnaire 3 asked them to select their present behavioral frequency as compared with their own past learning experiences in secondary schools; and in Questionnaire 4, subjects were told to tick the relevant frequency of a behavior by comparing its frequency of occurrence with that of other language skills (see Appendices A and B for details).

It should be noted that it was not the intention of the present study to prove the existence of the three dimensions of reference mentioned in the previous paragraph, as these reference systems were taken from student interviews in one of our previous studies (Wen, 1993). What we did aim at was to show in more general terms whether these references were distinguishable among themselves and from the general questionnaire. If yes, we would argue, it would make

all the difference to have a questionnaire with and another one without specific references correlated with the same criterion measure, say, learning outcome.

Procedures

Twenty items were first taken or adapted from Oxford (1990) and then translated into Chinese. Questionnaire directions were next drawn up and translated as well. Then 120 copies of the four questionnaires were printed and sent to Beijing University of Industry for administration.

Questionnaire 1 was first given to the target population by their English teachers during class hours. A week later, the same students were given Questionnaires 2, 3, and 4 at the same class period in order to save time. In hindsight, however, these questionnaires should have been presented at about one week intervals, since it turned out that a number of subjects got bored and either copied answers from one questionnaire to another or simply did not fill in Questionnaires 3 and 4. This resulted in the dwindling of subject numbers for the latter two questionnaires.

Ideally, we could have used a Latin Square design to control for maturation/time effect by manipulating the temporal order of questionnaire administration. However, for a small scale study of this kind, the laborious nature of such procedures rendered it impractical in obtaining cooperation from the subjects and the questionnaire administrators. And furthermore, we would argue that unlike in proficiency tests, time effect is relatively less of a problem in administering similar questionnaires several times.

Analyses

Correlation analysis was first done to see if the four questionnaires measured different things because of the change in reference system. Next, in order to know whether significantly different answers would result from the four questionnaires, data from the four questionnaires were subjected to repeated measures MANOVA analysis. Since post hoc multiple group comparisons for MANOVA are not readily available in SPSS/PC+, we used nonorthogonal CONTRAST procedures ("simple" and "repeated") under the TRANSFORM subcommand to locate significant differences among the questionnaires. Listwise deletion of missing data further reduced the sample size to 53-54 depending on the number of responses to individual items. Descriptive statistics were also obtained to show how exactly the same students differed in answering the questionnaires.

RESULTS

Do the Four Questionnaires Measure the Same Things?

It follows that if subjects think in different reference systems when a Questionnaire 1 type of measurement is presented to them, and that if the reference systems are indeed very far apart from each other, they will have significant validity implications for the measurement. To see how closely the three specific reference systems are related to each other and how they relate to the general questionnaire with no reference specifications, correlation analysis was performed. Results are presented in Appendix C.

Overall, correlation coefficients among the four questionnaires ranged from -.18 to .70. Questionnaires 2 and 3 were the most highly correlated, and Questionnaires 1 and 4 were the least correlated. Indeed the only two negative correlations (albeit insignificant) among all questionnaire items were between Questionnaires 1 and 4, and only two out of twenty items between these two questionnaires were significantly correlated. Of the 20 correlation coefficients between Questionnaires 1 and 2, 18 were significant, and 11 coefficients were significant between Questionnaires 1 and 3. In Questionnaires II and IV, 14 items were correlated significantly, and 18 significant coefficients were found between Questionnaires 3 and 4.

Do Subjects Answer the Four Questionnaires Differently?

Appendix D summarizes descriptive statistics and results from the repeated measures MANOVA analysis. *F* values and their probabilities indicate that out of the 20 items, 13 were significantly different at the .05 level, showing that students did answer the four questionnaires differently at least in 13 of the 20 items. In addition, a number of items (e.g., 2, 11, 17, 18) produced significant differences between one or two pairs of comparisons out of a total possibility of six.

To perform a general comparison of the four questionnaires, all 20 items were added up and divided by 20 to obtain the grand mean of a questionnaire for a single subject. Repeated measures MANOVA was again performed to see whether these four questionnaires in general differed from each other significantly. Results in Table 3 [$F(3,156)=10.48$, $p<.001$] did reveal significant differences amongst the four questionnaires. Moreover, subsequent contrast procedures locate specific differences between Questionnaires 1 and 2, 1 and 3, 1 and 4, and between Questionnaires 2 and 3.

Table 1.

General Comparisons Among the Four Questionnaires

Q'naire	*N*	Grand Mean	*SD*	Grand Order	*F*	*df*	*p*	Contrast Q1	Q2	Q3	Q4
1	92	2.556	.408	1	10.48	3,156	.000				
2	95	2.646	.496	2				*			
3	80	2.802	.574	4				*	*		
4	55	2.754	.466	3				*			

* $p<.05$

Tables 1 and 2 also show the ordering of the four questionnaires. A *1* means that the questionnaire obtained the lowest mean score, *2* the second lowest, *3* the one that ranked third, and *4* the one that ranked the highest. In general, Questionnaire 1 ranked the lowest (Grand Mean=2.556), Questionnaire 2 the second lowest (Grand Mean=2.646), Questionnaire 4 the second highest (Grand Mean=2.754) and Questionnaire 3 the highest (Grand Mean=2.802).

Table 2.

Ordering of the Four Questionnaires in the Twenty Items

Q'naire	First		Second		Third		Fourth		Grand Order
	Frq	%	Frq	%	Frq	%	Frq	%	
1	11	55	6	30	0	0	3	15	1
2	5	25	10	50	5	25	0	0	2
3	0	0	3	15	7	35	10	50	4
4	4	20	1	5	8	40	7	35	3

More specifically, Questionnaire 1, the one that did not specify any system of reference, occurred in 11 items (55%) with the lowest mean score, and it occurred in 6 items (30%) as the second lowest, and in only 3 items (15%) as the highest. This would probably mean that when no reference is specified, as is often the case in questionnaire research, subjects would be more cautious in responding to the Likert-scale. Likewise, Questionnaire 2, the one that asked the subjects to rank-order themselves among their classmates, ranked the second lowest, possibly indicating again (cf. Zax & Takahashi, 1967) the modesty of the Asians, in this case, the Chinese learners. Clearly, Questionnaire 3 ranked the highest, meaning that the subjects were applying more of the activities specified in the questionnaires than they had done in their secondary schools. Understandably, Questionnaire 4 did not have a mode as the other three questionnaires did, as it asked the subjects to respond to an item by comparing with what they did with other language skills within the same questionnaire.

DISCUSSION

On the whole, this study has focused on one problem with the Likert-scale in language learning strategy research, namely, the ambiguities of the scale caused by ambiguities of reference. Three specific dimensions of reference taken from student interviews in Wen (1993) have been compared with the usual type of questionnaire as well as among themselves. It has been shown that the same statements can elicit very different things when understood in different reference systems (e.g., Questionnaires 1 and 4). In general, the Chinese subjects in this study tended to be cautious in responding to the Likert-scale on learning activities when presented with a general questionnaire with no explicit reference. When clear references were provided, responses to the same questions were mostly different from those elicited through the general questionnaire.

Despite methodological limitations of this study as mentioned earlier, the previous results generally confirm our original conception that different reference systems not only exist but also vary significantly among themselves. Moreover, questionnaires that specified these references differed significantly from the questionnaire that did not indicate any reference. Our message is by now loud and clear: questionnaire research on learning behaviors that uses Likert scaling as its only means of data elicitation may well be dangerously inadequate and unreliable. For one thing, the problem of reference ambiguity must be seriously considered before any conclusion is reached from such a study.

Implications of the present study go beyond Likert scaling *per se*. With only a handful of exceptions in recent years (e.g., Naiman et al., 1978; Stevick, 1989; Vann & Abraham, 1990), positivism still prevails over the field of SLA. Numbers are normally hallowed and speak louder than ideas. However, as Shipman (1988) and so many others rightly point out, quantitative social and behavioral sciences are not at all free of subjective biases, academic dogmatism, and even malicious fraud. Apart from its almost congenital superficiality, we would posit that quantitative research suffers most from its abuses. Using numbers the wrong way is worse than not using them at all, especially in areas like language learning strategies, where, as Gu (1992) noted, the strategies concerned can hardly be reduced to frequency counts, and where the qualitative alternative has been producing very revealing results. Fortunately, researchers in SLA are beginning to ponder over methodological triangulation (e.g., Cohen, 1984; Grotjahn, 1991) so as to counterbalance any limitations of one particular method.

Notwithstanding the alarm we have raised so far about the use of quantitative methods, and particularly about the Likert scale in SLA research, we do not wish to be understood as being against either the quantitative research tradition in general or the Likert-scale in particular. On the contrary, we view quantitative research as an invaluable and even the best possible approach to finding patterns, establishing relations (especially causations), and making predictions. We would thus be more than delighted if this study has shed some light on one of the many pitfalls in empirical research.

We believe that our warnings about the possible abuses of the Likert-scale in this study are of particular significance to the Chinese EFL research arena in view of our marked lack of quantitative expertise (Meng, 1993). Amid acute cries in recent years for statistical literacy among Chinese EFL researchers (e.g., Liu, 1992), we have to be aware at the same time of the inadequacies of quantitative research, and even of the much needed balance of emphasis between the two ever-lasting epistemological ends: positivism and rationalism. Simply put, a basic philosophy of research might be just as badly needed as the methods and techniques of research, let alone statistics. Knowing the legitimate scientific alternatives to quantitative research, and knowing why and when to play with numbers are at least as important as, if not more important than, knowing how to conduct empirical research.

Future research into language learning behaviors employing the Likert-scale should be cautioned against its ambiguities. Semantic implications of *oftenness* as well as subjective references as shown in this study should be made as explicit and unequivocal as possible. For example, *about two hours a day every day* is much clearer than *50 percent of the time*. Furthermore, we strongly recommend the use of methodological triangulation in language learning strategy research, e.g., following up a major questionnaire study with in-depth interviews and/or task-based think-aloud procedures, or vice versa, so as to avoid the aforementioned problems.

CONCLUSION

A Chinese fable (documented in *Lushi Chunqiu: Chajin*) goes as follows:

> A man from the state of Chu was travelling in a boat when his sword fell into the river. He instantly drew out his dagger and cut a mark on one side of the boat and said to himself: This is where my sword fell. After the boat finally pulled ashore, he jumped into the water from the marked point of the boat and searched in vain for his sword. Amongst whatever morals we can draw from this fable, one is clear: When

trying to get something, we need to get it the right way, or we risk not getting it at all. After all, the Chuvian has been laughed at for over two millennia, how long are we as researchers prepared to be laughed at for re-searching for the sword of learning strategies?

REFERENCES

Busch, M. (1993). Research issues: Using Likert scales in L2 research. *TESOL Quarterly, 27*(4), 733-736.

Cohen, A. D. (1984). Studying second-language learning strategies: How do we get the information? *Applied Linguistics, 5*(2), 101-112.

Dunn-Rankin, P. (1983). *Scaling methods*. Hillsdale, NJ: Erlbaum.

Grotjahn, R. (1991). The research programme subjective theories: A new approach in second language research. *Studies in Second Language Acquisition , 13*(2), 187-214.

Gu, Yongqi. (1992). *Robin Hood in SLA: What has the learning strategy researcher taught us?* Unpublished manuscript. Department of Education, University of Hong Kong.

Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of CrossCultural Psychology, 20* (3), 296-309.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* . Englewood Cliffs, NJ: Prentice Hall.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1-55.

Liu, R. (1992). ELT in China 1992 yantaohui bufen lunwen jianjie [A brief introduction to the key articles presented at the ELT in China 1992 Conference]. *Waiyu Jiaoxue yu Yanjiu, 90*(2), 3-13.

Meng, Yue. (1993). Muqian woguo yingyong yuyanxue yanjiu fangfa de diaocha yu fenxi [A survey analysis of applied linguistics research methods in China]. *Xiandai Waiyu, 59*, 1-6.

Naiman, N., Frohlich, M., Stern, H. H., & Todesco, A. (1978). *The good language learner*. Toronto, Ontario: Institute for Studies in Education.

Oxford, R.L. (1990). *Language learning strategies: What every teacher should know* . New York: Newbury House.

Oxford, R. L., & Nyikos, M. (1989). Variables affecting choice of language learning strategies by university students. *Modern Language Journal, 73*(2), 291-300.

Oxford, R. L., Nyikos, M., & Ehrman, M. (1988). Vive la difference? Reflections on sex differences in use of language learning strategies. *Foreign Language Annals, 21*(4), 321-329.

Politzer, R. L., & McGroarty, M. (1985). An exploratory study of learning behaviors and their relationship to gains in linguistic and communicative competence. *TESOL Quarterly, 19*, 103-123.

Shaw, M. E., & Wright, J. M. (1967). *Scales for the measurement of attitudes* . New York: McGraw-Hill.

Shipman, M. (1988). *The limitations of social research* (3rd ed.). London: Longman.

Stevick, E. W. (1989). *Success with foreign languages: Seven who achieved it and what worked for them*. London: Prentice Hall.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*, 529-554.

Triandis, H. C., & Triandis, L. M. (1962). A cross-cultural study of social distance. *Psychological Monographs*, *76*(21).

Turner, J. (1993). Research issues: Using Likert scales in L2 research. *TESOL Quarterly*, *27*(4), 736-739.

Vann, R. J., & Abraham, R. G. (1990). Strategies of unsuccessful language learners. *TESOL Quarterly*, *24*(2), 177-198.

Wen, Q. (1993). *Advanced level English language learning in China: The relationship of modifiable learner variables to learning outcomes*. Unpublished doctoral dissertation. Department of Education, University of Hong Kong, Hong Kong.

Wright, G. N., Phillips, L. D., Whalley, P. C., Choo, G. T., Ng, K. O., Tan, I., & Wisudha, A. (1978). Cultural differences in probabilistic thinking. *Journal of Cross-Cultural Psychology*, *9*, 285-299.

Zax, M., & Takahashi, S. (1967). Cultural influences on response style: Comparisons of Japanese and American college students. *Journal of Social Psychology*, *71*, 3-10.

An earlier version of this paper was presented at the 1993 Guilin ELT International Conference. The authors would like to thank Dr. Robert Keith Johnson, Mr. Raymond Lam, and the three anonymous reviewers for their helpful comments on that version. They also extend special thanks to Ms. Chen Xiaoming and her students at Beijing University of Industry for their kind cooperation during data collection.

Yongqi Gu is currently completing his Ph.D. in the Department of Education at the University of Hong Kong. His main research interests include learning processes and learning strategies, vocabulary, and research methods.

Qiufang Wen obtained her Ph.D. in 1993 from the Department of Education at the University of Hong Kong. She is now an associate professor at the School of Foreign Languages of Nanjing University.

Danxia Wu is a lecturer of English at the Central Academy of Fine Arts, Beijing. She has extensive experience teaching English to non-English majors in China.

APPENDIX A

Directions of the Four Questionnaires

A. The Five -Point Scale From Oxford (1990)

1=Never or almost never true of me

2=Rarely true of me

3=Somewhat true of me

4=often true of me

5=Always or almost always true of me

B. Directions of the Four Questionnaires

Questionnaire 1

This questionnaire contains twenty statements about English learning. Please read each statement. Write your response (1, 2, 3, 4, or 5) that tells how true of you the statement is in the blanks preceding each statement.

Questionnaire 2

This questionnaire contains twenty statements about English learning. Please read each statement. Write in the blanks preceding each statement your response (1, 2, 3, 4, or 5) that tells how true of you the statement is **as compared to your classmates**. E.g., A choice of 2 means most of your classmates do this more often than you do.

Questionnaire 3

This questionnaire contains twenty statements about English learning. Please read each statement. Write in the blanks preceding each statement your response (1, 2, 3, 4, or 5) that tells how true of you the statement is **as compared to your past learning experience in secondary schools**. E.g., 4 means you often do this now, but you didn't do it as often in the past.

Questionnaire 4

This questionnaire contains twenty statements about English learning. Please read each statement. Write in the blanks preceding each statement your response (1, 2, 3, 4, or 5) that tells how true of you the statement is **as compared to what you do to improve other language skills**. E.g., If you choose 5 for item number 1, you always look for opportunities to LISTEN to English, but perhaps not as often in finding opportunities to speak, read, and write.

APPENDIX B

The Twenty Items in the Four Questionnaires

Listening

1. I look for opportunities to listen to English.
2. While listening, I make guesses to understand unfamiliar English words.
3. I listen for pleasure in English.
4. I make summaries of information that I hear in English.
5. I try to guess what the other person will say next in English based on what has been said so far.

Speaking

6. I look for opportunities to speak as much English as possible.
7. I try to use newly learned words and expressions when I speak English.
8. When I speak, I make up new words if I do not know the right ones in English.
9. If I can't think of an English word when I speak, I use a word or phrase that means the same thing.
10. I notice my own mistakes and use that information to help me improve my spoken English.

Reading

11. I look for opportunities to read as much as possible in English.
12. When I read English, I make guesses to understand unfamiliar English words.
13. I read for pleasure in English.
14. I make summaries of information that I read in English.
15. When I read, I try to guess what the author will write next based on what has been written so far.

Writing

16. I try to write everything I can in English.
17. I try to use newly learned words and expressions when I write in English.
18. When I write, I make up new words if I do not know the right ones in English.
19. If I can't think of an English word when I write, I use a word or phrase that means the same thing.
20. I notice my own mistakes and use that information to help me write better English.

APPENDIX C

Correlations Among the Four Questionnaires

	Questionnaire 1	Questionnaire 2	Questionnaire 3	Questionnaire 4
Item 1				
Q1				
Q2	.4677**			
Q3	.3229*	.3933**		
Q4	.3337*	.3273*	.6958**	
Item 2				
Q1				
Q2	.2352			
Q3	.0240	.3369*		
Q4	-.1779	.2093	.4551**	
Item 3				
Q1				
Q2	.2882*			
Q3	.2773*	.6056**		
Q4	.2186	.1518	.3563*	
Item 4				
Q1				
Q2	.4938**			
Q3	.2017	.3711**		
Q4	.1463	.4354**	.4786**	
Item 5				
Q1				
Q2	.2713*			
Q3	.1418	.3283*		
Q4	.0702	.3361*	.5088**	
Item 6				
Q1				
Q2	.3911**			
Q3	.2827*	.3869**		
Q4	.1076	.3056	.6169**	

	Questionnaire 1	Questionnaire 2	Questionnaire 3	Questionnaire 4
Item 7				
Q1				
Q2	.3136*			
Q3	.2777*	.5053**		
Q4	.1259	.3801*	.4579**	
Item 8				
Q1				
Q2	.3242**			
Q3	.3751**	.5204**		
Q4	.3019	.2889	.3596*	
Item 9				
Q1				
Q2	.2412			
Q3	.1546	.5118**		
Q4	-.0872	.3964*	.3477*	
Item 10				
Q1				
Q2	.2550*			
Q3	.1692	.5794**		
Q4	.0353	.3922*	.4863**	
Item 11				
Q1				
Q2	.4261**			
Q3	.2501	.4642**		
Q4	.1771	.1573	.3683*	
Item 12				
Q1				
Q2	.3435**			
Q3	.2190	.3783**		
Q4	.1998	.0484	.4433**	
Item 13				
Q1				
Q2	.4431**			
Q3	.4165**	.6255**		
Q4	.1642	.5339**	.4346**	

	Questionnaire 1	Questionnaire 2	Questionnaire 3	Questionnaire 4
Item 14				
Q1				
Q2	.3931**			
Q3	.3287*	.4428**		
Q4	.2027	.3986*	.2767	
Item 15				
Q1				
Q2	.4421**			
Q3	.3182*	.4174**		
Q4	.3826*	.5041**	.1810	
Item 16				
Q1				
Q2	.5165**			
Q3	.3007*	.4874**		
Q4	.1307	.4517**	.3641*	
Item 17				
Q1				
Q2	.4101**			
Q3	.1273	.5116**		
Q4	.1903	.3699*	.3955*	
Item 18				
Q1				
Q2	.3132*			
Q3	.3502**	.6174**		
Q4	.2473	.3531*	.5311**	
Item 19				
Q1				
Q2	.4007**			
Q3	.3266*	.5062**		
Q4	.2719	.4758**	.4740**	
Item 20				
Q1				
Q2	.3556**			
Q3	.1958	.6957**		
Q4	.1778	.5017**	.5619**	

* $p<.01$ ** $p<.001$

APPENDIX D

Descriptive Statistics, Repeated Measures MANOVA and Multiple Contrasts

Item	Q'naire	*N*	Mean	*SD*	*F*	*df*	*p*	Contrast			
								Q1	Q2	Q3	Q4
1	1	92	2.76	.70	6.08	3,156	.001				
	2	95	2.62	.76							
	3	80	3.08	1.09				*	*		
	4	56	3.00	.89					*		
2	1	92	2.92	1.00	1.59	3,159	.193				
	2	95	2.97	.88							
	3	81	3.16	.86							
	4	56	2.86	.88						*	
3	1	92	1.99	.85	15.16	3,159	.000				
	2	95	2.27	.99				*			
	3	81	2.75	1.03				*	*		
	4	56	2.80	1.03				*	*		
4	1	92	2.11	.92	13.23	3,159	.000				
	2	95	2.49	.91				*			
	3	81	2.74	1.07				*	*		
	4	56	2.73	.88				*			
5	1	92	2.55	1.05	2.74	3,159	.045				
	2	95	2.77	1.04							
	3	81	2.91	1.04				*			
	4	56	2.89	.89				*			
6	1	92	1.70	.62	19.48	3,156	.000				
	2	95	2.17	.74				*			
	3	81	2.51	.85				*	*		
	4	55	2.36	.87				*	*		
7	1	92	2.11	.86	10.47	3,159	.000				
	2	95	2.49	.85				*			
	3	81	2.59	.89				*			
	4	56	2.59	.85				*			

Item	Q'naire	N	Mean	SD	F	df	p	Contrast			
								Q1	Q2	Q3	Q4
8	1	92	2.28	1.07	1.05	3,159	.373				
	2	95	2.49	.99							
	3	81	2.57	.92							
	4	56	2.54	.95							
9	1	92	3.22	.92	2.60	3,159	.054				
	2	95	3.16	.80							
	3	81	3.00	.96							
	4	56	2.80	1.02				*	*		
10	1	92	2.11	.87	7.71	3,159	.000				
	2	95	2.44	.91				*			
	3	81	2.68	.97				*	*		
	4	56	2.30	.78						*	
11	1	92	3.08	.84	1.71	3,156	.167				
	2	95	2.84	.89				*			
	3	81	3.09	1.01							
	4	55	3.09	.99							
12	1	92	3.55	.72	8.82	3,159	.000				
	2	95	3.28	.85				*			
	3	81	3.02	.87				*	*		
	4	56	3.00	.91				*	*		
13	1	92	2.30	.89	11.47	3,159	.000				
	2	95	2.45	.81							
	3	81	2.56	.88				*			
	4	56	2.98	1.02				*	*	*	
14	1	92	2.74	.97	1.61	3,159	.188				
	2	95	2.71	.91							
	3	81	2.81	.96							
	4	56	2.88	.81							
15	1	92	2.51	.93	3.28	3,159	.022				
	2	95	2.60	.88							
	3	81	2.81	.91				*			
	4	56	2.95	.86				*	*		
16	1	92	2.38	.98	4.63	3,156	.004				
	2	95	2.38	.95							
	3	80	2.70	.95				*	*		
	4	56	2.59	.87				*			

Item	Q'naire	*N*	Mean	*SD*	*F*	*df*	*p*	Contrast			
								Q1	Q2	Q3	Q4
17	1	92	2.52	.83	2.03	3,159	.112				
	2	95	2.72	.85				*			
	3	81	2.78	.85							
	4	56	2.79	.78				*			
18	1	92	2.40	.97	1.67	3,159	.176				
	2	95	2.40	.89							
	3	81	2.54	.88					*		
	4	56	2.63	.93							
19	1	92	3.43	.73	8.64	3,159	.000				
	2	95	3.15	.81				*			
	3	81	3.04	.90				*			
	4	56	2.82	.92				*	*		
20	1	92	2.45	.86	1.89	3,159	.134				
	2	95	2.51	.85							
	3	81	2.68	.97							
	4	56	2.52	.81							

* $p<.05$